A general class of estimators of the extreme value index

Holger Drees*

Mathematisches Institut, Universität zu Köln, Weyertal 86-90, 50931 Köln, Germany

Abstract

We consider the class of estimators of the extreme value index $\beta$ that can be represented as a scale invariant functional $T$ applied to the empirical tail quantile function $Q_n$. From an approximation of $Q_n$ first asymptotic normality of $T(Q_n)$ is derived under quite natural smoothness conditions on $T$ if $\beta$ is positive. As a consequence, a widely applicable method for the construction of estimators with a prescribed asymptotic behavior is introduced. If $\beta \leq 0$ then either $T$ must be location invariant or it has to satisfy a certain regularity condition on a neighborhood of a constant function to ensure asymptotic normality. It turns out that in this situation location invariant estimators are clearly preferable. © 1998 Elsevier Science B.V.

AMS classification: primary 62G05; secondary 62G30, 60F17

Keywords: Extreme value distribution; Extreme value index; Statistical tail functional; Empirical tail quantile function; Hadamard differentiability; Location invariance; Adaptive estimator

1. Introduction

Suppose $n$ i.i.d. random variables (r.v.'s) $X_i$, $1 \leq i \leq n$, with common distribution function (d.f.) $F$ are observed. It is assumed that $F$ belongs to the weak domain of attraction of some unknown extreme value d.f. $G$ (in short, $F \in D(G)$), that is,

$$U(a_{n}^{-1}\left(\max_{1 \leq i \leq n} X_i - b_{n}\right)) \rightarrow G \quad \text{weakly (1.1)}$$

for some normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$.

Since the famous work of Gnedenko (1943) it is well known that up to a scale and location parameter a nondegenerate limiting d.f. has to be of the form

$$G_{\beta}(x) = \exp(-((1 + \beta x)^{-1/\beta})) \quad 1 + \beta x > 0, \quad \beta \in \mathbb{R},$$

which is interpreted as $\exp(-e^{-x})$ if $\beta = 0$.

* E-mail: hdrees@mi.uni-koeln.de.
The problem of estimating the so-called extreme value index $\beta$, which determines the behavior of the underlying d.f. $F$ in its upper tail, has received much attention in the last twenty years; see, e.g., Hill (1975), Pickands (1975), Hall and Welsh (1984, 1985), Csörgő et al. (1985), Smith (1987), Dekkers et al. (1989) and Drees (1995a, b, 1996). An extensive motivation of this estimation problem can be found in Galambos (1978); more recently de Haan and Rootzén (1993) demonstrated how to use an estimate of $\beta$ to construct estimators of extreme quantiles.

In view of (1.1) it is evident that an estimator of $\beta$ should utilize only large observations. Essentially, there are two possible interpretations of what is to be considered as large in this context. The Gumbel or annual maxima method utilizes the maxima of disjointed subsamples. Such an approach suggests itself, for example, if the data are collected over a period that splits up into several subperiods in a natural way. On the other hand, the method seems quite arbitrary if there is no such natural partition of the sample.

For that reason, since the paper of Pickands (1975) more attention has been paid to estimators $\hat{\beta}_n$ that are based on a certain (deterministic or random) number, say $k_n + 1$, of upper order statistics $X_{n-i+1:n}$, $1 \leq i \leq k_n + 1$. Then one can find a functional $T_n$ such that

$$\hat{\beta}_n = T_n(Q_n),$$

where the empirical tail quantile function (q.f.) $Q_n$ is defined as

$$Q_n(t) := F_n^{-1}\left(1 - \frac{k_n}{n}t\right) = X_{n-[k_n:t]:n}, \quad t \in [0,1].$$

If one wants to investigate the asymptotic behavior of such a sequence of estimators, then clearly, one has to assume that the sequence of functionals $T_n$ converges to a limiting functional $T$ in some sense. Thus, it seems reasonable to assume that the sequence is even constant, that is,

$$\hat{\beta}_n = T(Q_n) \quad (1.2)$$

for a functional $T$ and all $n \in \mathbb{N}$. In fact, almost all known estimators of the extreme value index that are based on largest order statistics are of this type, for example, the Hill estimator, the Pickands estimator and the kernel estimators considered by Csörgő et al. (1985). Observe that estimators of the form (1.2) can be regarded as an extreme value counterpart of the classical statistical functionals (see Fernholz, 1983). Therefore, they will be addressed as statistical tail functionals in the sequel.

In the following, assume that the number of order statistics used for the estimation constitutes a deterministic intermediate sequence, i.e., $k_n \rightarrow \infty$ and $k_n/n \rightarrow 0$. Note,
however, that all results established below carry over to estimators that are based on the exceedances over a given deterministic threshold if the average number of these exceedances equals $k_n$.

In a foregoing paper (Drees, 1995b) we investigated the asymptotic behavior of statistical tail functionals $T(Q_n)$ under the assumption that $T$ is scale and location invariant. Indeed there are mathematical, as well as practical reasons to require such invariance properties. Since an affine transformation of the r.v.'s $X_i$ merely leads to a change of the normalizing constants $a_n$ and $b_n$, it influences neither the extreme value index nor the accuracy of the approximation (1.1). Moreover, in practice, it is reasonable to demand that a change of the unit of measure, which implies a scale transformation of the data, will not change the estimate of $\beta$. Finally, in some applications, the observations (e.g., sea levels or temperatures) depend on an arbitrarily chosen zero-point, which equally should not affect the estimator.

On the other hand, while all the well-known estimators of the extreme value index are scale invariant, there are some that are not invariant under a shift of the data. Certainly, the best-known estimator of this type is the Hill estimator, which is only consistent for nonnegative $\beta$, but there are also estimators that work for all real extreme value indices, e.g., the moment estimator proposed by Dekkers et al. (1989).

Therefore, in the present paper we aim at extending the results given in Drees (1995b) to the general class of statistical tail functionals that are scale but not location invariant. It will turn out that there are only limited changes necessary if $\beta>0$, but that the situation alters completely if the extreme value index is negative. Of course, this is due to the fact that a shift of the maximum of the observations leads to a different limit if the underlying d.f. has a finite right endpoint whereas, roughly speaking, $(X_{n:n} + \mu)/X_{n:n} = O(n^{-\beta})$ if $\beta$ is positive. Hence, we have to deal with these two cases separately.

In Section 2 an approximation of the empirical tail q.f., which was established in Drees (1995b), is prepared for an easy application in the absence of location invariance. As a consequence, in Section 3 we obtain asymptotic normality under a differentiability condition on the functional $T$ that is comparable to the one imposed in our foregoing paper. Two examples, one of which concerns Hill's estimator, demonstrate the applicability of the result. In particular, a general method is proposed to design estimators of the extreme value index with prescribed asymptotic behavior. The section closes with a discussion of the asymptotic efficiency of the estimators under consideration, particularly in comparison with the location invariant statistical functionals. In Section 4 we go on to the case where the extreme value index is not positive. It turns out that a condition on the local behavior of the functional $T$ in a neighborhood of the constant function 1 that seems less natural than the differentiability condition in the case $\beta>0$ is necessary to obtain asymptotic normality. However, there are examples of this type of statistical functionals including the moment estimator of Dekkers et al. (1989). An examination of the limiting normal distribution leads to the conclusion that at least asymptotically location invariant estimators are preferable if the extreme value index is expected to be negative.
2. A limit theorem for the empirical tail quantile function

If one wants to derive an approximation of the empirical tail q.f. that is based on the convergence (1.1), then first, one has to quantify the rate of that convergence. This can be done by means of the so-called second-order conditions. In literature, a lot of such conditions can be found, see, e.g., Hall (1982), Csörgő et al. (1985), Smith (1987), Reiss (1989) and Dekkers and de Haan (1989, 1993). However, due to its simplicity and generality, the following expansion of the underlying q.f., which was extensively studied by de Haan and Stadtmüller (1996), seems most attractive.

**Condition 1.** There exist measurable, locally bounded functions \( a, \Phi : (0, 1) \to (0, \infty) \) and \( \Psi : (0, \infty) \to \mathbb{R} \) such that

\[
\frac{F^{-1}(1- tx) - F^{-1}(1- t)}{a(t)} = \frac{x^{-\beta} - 1}{\beta} + \Phi(t)\Psi(x) + R(t,x)
\]

for all \( t \in (0, 1) \) and \( x > 0 \). (By convention, \( (x^{-\beta} - 1)/\beta := - \log(x) \) if \( \beta = 0 \).) Here

(i) \( \Psi \equiv 0 \) and \( R(t,x) = o(1) \) as \( t \downarrow 0 \) for all \( x > 0 \) or

(ii) \( x \to \Psi(x)/(x^{-\beta} - 1) \) is not constant, \( \Phi(t) = o(1) \) and \( R(t,x) = o(\Phi(t)) \) as \( t \downarrow 0 \) for all \( x > 0 \).

Note that Condition 1(i) is equivalent to our basic assumption \( F \in D(G_\beta) \) (de Haan, 1984, Theorem 1). The stronger Condition 1(ii) implies that \( \Phi \) is \( \delta \)-varying at 0 for some \( \delta > 0 \), i.e., \( \Phi(tx)/\Phi(t) \to x^\delta \) as \( t \downarrow 0 \), and by a suitable choice of the functions \( a \) and \( \Phi \) one can achieve that

\[
\Psi(x) = c_\Psi \cdot \begin{cases} 
  x^{\delta-\beta} - 1 & \beta \neq \delta > 0, \\
  \log(x) & \beta = \delta > 0, \\
  x^{-\beta} \log(x) & \beta \neq \delta = 0, \\
  \log^2(x) & \beta = \delta = 0.
\end{cases}
\]

For a detailed discussion of Condition 1(ii) we refer to de Haan and Stadtmüller (1996) and Drees (1995b).

Next we need some notation. Let \( h \) be some function belonging to

\[ \mathcal{H} := \left\{ h : [0, 1] \to [0, \infty) \mid h \text{ continuous, } \lim_{t \downarrow 0} h(t) \left( \frac{\log \log(1/t)}{t} \right)^{1/2} = 0 \right\}. \]

Then the function space

\[ D_{\beta, h} := \left\{ z : [0, 1] \to \mathbb{R} \mid \lim_{t \downarrow 0} t^\beta h(t)z(t) = 0, (t^\beta h(t)z(t))_{t \in [0, 1]} \in D[0, 1] \right\} \]

is equipped with the weighted supremum seminorm

\[ \|z\|_{\beta, h} := \sup_{t \in [0, 1]} t^\beta h(t)|z(t)|. \]
Furthermore, we introduce the functions

\[
    z_\beta(t) := \begin{cases} 
    t^{-\beta} & \beta > 0, \\
    -\log(t) & \beta = 0, \\
    -t^{-\beta} & \beta < 0.
    \end{cases}
\]

\[
    \overline{\Psi} := \Psi + c_\Psi \cdot 1_{\{\beta \neq 0\}}, \quad \Xi(t) := \begin{cases} 
    \beta F^{-1}(1 - t)\alpha(t) - 1 - \beta c_\psi \Phi(t) \cdot 1_{\{\beta \neq 0\}} & \beta > 0, \\
    F^{-1}(1 - t)\alpha(t) - c_\psi \Phi(t) \cdot 1_{\{\beta > 0\}} & \beta = 0, \\
    c_\alpha |\beta|/\alpha(t) & \beta < 0.
    \end{cases}
\]

where \( c_\alpha = F^{-1}(1) \) denotes the right endpoint of \( F \). Finally, let

\[
    A_n := |\beta|(X_{n:n} - \omega)/\alpha(k_n/n) \cdot 1_{\{\beta < -1/2\}}, \\
    e_\beta := \begin{cases} 
    |\beta| & \beta \neq 0, \\
    1 & \text{if } \beta = 0.
    \end{cases}
\]

and define a Gaussian process by

\[
    V_\beta(t) := e_\beta t^{-\beta+1}W(t)
\]

where \( W \) denotes a standard Brownian motion.

Then the following limit theorem for the empirical tail q.f. is a reformulation of Corollary 2.1 of Drees (1995b).

**Theorem 2.1.** There are constants \( c_n \) such that the following assertions hold for all \( h \in H \):

(i) If Condition 1(i) and

\[
    \sup_{x \in [0, 1+\varepsilon]} x^{\beta+1/2} |R(k_n/n, x)| = o(k_n^{-1/2})
\]

for some \( \varepsilon > 0 \), then

\[
    k_n \left( \frac{Q_n}{c_n} - (z_\beta + \Xi(k_n/n) + A_n) \right) \rightarrow V_\beta,
\]

weakly in \( D_{\beta,h} \).

(ii) Under Condition 1(ii) one has

\[
    \Phi(k_n/n) = O(k_n^{-1/2}) \Rightarrow k_n^{1/2} \left( \frac{Q_n}{c_n} - (z_\beta + \Xi(k_n/n) + A_n + e_\beta \Phi(k_n/n)\overline{\Psi}) \right) \rightarrow V_\beta
\]
and
\[ k_n^{-1/2} = o(\Phi(k_n/n)) \]
\[ \Rightarrow \Phi(k_n/n)^{-1} \left( \frac{Q_n}{c_n} - (z_\beta + \Xi(k_n/n) + \Delta_n + e_\beta \Phi(k_n/n) \bar{\Phi}) \right) \rightarrow 0, \]
weakly in \( D_{\beta,h} \).

**Proof.** In Drees (1995b) it was shown that if Condition 1(ii) is satisfied and \( \Phi(k_n/n) = O(k_n^{-1/2}) \), then
\[ k_n^{1/2} \left( \frac{Q_n - \tilde{D}_n}{a(k_n/n)} - \left( \frac{z_\beta}{e_\beta} + \Phi(k_n/n) \bar{\Phi} \right) \right) \rightarrow \frac{V_\beta}{e_\beta}, \]
weakly in \( D_{\beta,h} \) where
\[ \tilde{D}_n = \begin{cases} F^{-1}(1 - k_n/n) - a(k_n/n)(1_{\beta\neq0}/\beta + c_\Psi \Phi(k_n/n) \cdot 1_{\beta\neq0}) & \text{if } \beta \geq -\frac{1}{2}, \\ X_n & \text{if } \beta < -\frac{1}{2}. \end{cases} \]

Hence, the assertion is immediate in this case if one takes into account that by Drees (1995b), Lemma 2.1, for \(-\frac{1}{2} \leq \beta < 0\)
\[ \frac{\omega - F^{-1}(1 - k_n/n)}{a(k_n/n)} + \frac{1}{\beta} + c_\Psi \Phi(k_n/n) 1_{\beta > 0} = o(\Phi(k_n/n)) = o(k_n^{-1/2}). \]
The other cases can be treated similarly. \( \square \)

Recall that according to Drees (1995b), Lemma 2.1, for a suitably chosen normalizing function \( a \), the left-hand side of (2.1) tends to 0 for any intermediate sequence \( k_n \) so that condition (2.1) is satisfied for some sequence. Note that though \( \Delta_n = o_p(k_n^{-1/2}) \) it has to be taken into account to ensure that the process under consideration belongs to \( D_{\beta,h} \) if \( \beta < -\frac{1}{2} \). In the case \( \beta > 0 \) a somewhat related result was established by Einmahl (1992).

Observe that \( \Xi \) depends on the parameter of interest but also on a scale parameter and the right endpoint \( \omega \) so that it seems impossible to extract information about \( \beta \) from \( \Xi(k_n/n) \). Therefore, we consider it as an additional source of error, which should vanish as the sample size increases. Since \( \Xi(t) \rightarrow 0 \) if \( \beta > 0 \) but \( |\Xi(t)| \rightarrow \infty \) if \( \beta \leq 0 \) and \( \omega \neq 0 \), these two cases have to be treated separately.

3. Asymptotic normality in the case \( \beta > 0 \)

While it is obvious that the constant term \( \Xi(k_n/n) \) in the approximation of \( Q_n \) does not influence location invariant estimators, at first glance it is not clear whether or not asymptotically this also holds true for statistical tail functionals without this invariance property. It will turn out that both cases are possible if merely Condition 1(i) holds.
Under Condition 1(ii), in general, the term $\Xi(k_{n}/n)$ is negligible if $\beta > 0$ and dominates $\Phi(k_{n}/n)$ otherwise. However, there is one exception to this rule. De Haan and Stadtmüller (1996) proved in their Theorem 2 that Condition 1(ii) holds with $\beta < \delta$ if and only if for some constants $d_{1} > 0$ and $d_{2} \in \mathbb{R}$

$$g(t) := F^{-1}(1 - t) - d_{1}t^{-\beta} - d_{2}$$

is $(\delta - \beta)$-varying at 0. In this case $\Xi(k_{n}/n)$ is negligible if and only if $d_{2} = 0$. So in contrast to the situation in Drees (1995b), the conditions on $k_{n}$ must also take into account the rate of convergence of $\Xi(k_{n}/n)$.

**Condition 2.** Assume that $(k_{n})_{n \in \mathbb{N}}$ is an intermediate sequence satisfying

(i) $\max(k_{n}^{1/2}, \Xi(k_{n}/n)^{-1}) \sup_{X \in (0,1]} x^{\beta+1/2} |R(k_{n}/n,x)| \rightarrow 0$ for some $\varepsilon > 0$ and $k_{n}^{1/2} \Xi(k_{n}/n) \rightarrow \rho \in [0, \infty]$ or

(ii) $k_{n}^{1/2} \Phi(k_{n}/n) \rightarrow \lambda \in [0, \infty]$.

With the exception of the location invariance the smoothness conditions on $T$, which is defined on $D_{\beta,h}$, are essentially the same as in the foregoing paper:

**Condition 3.** For $\beta > 0$ and some $h \in \mathcal{H}$ the functional $T : D_{\beta,h} \rightarrow \mathbb{R}$ fulfills

(i) $T$ is $\mathcal{B}(D_{\beta,h}), \mathcal{B}(\mathbb{R})$-measurable,

(ii) $T(az) = T(z)$ \quad $\forall z \in D_{\beta,h}, \ a > 0$,

(iii) $T(z_{\beta}) = \beta$,

(iv) $T$ is Hadamard differentiable tangentially to $C_{\beta,h}$ at $z_{\beta}$ with derivative $T_{\beta}'$.

Here $\mathcal{B}(X)$ denotes the Borel-$\sigma$-field of a metric space $X$ and $C_{\beta,h}$ is the space of continuous functions belonging to $D_{\beta,h}$.

**Remark.** Sometimes it is useful to specify $T$ only on some scale invariant subset of $D_{\beta,h}$ that contains $z_{\beta}$ and the paths of $Q_{n}$ with a probability converging to 1. Then it is sufficient to require Condition 3 (ii) and (iv) on this subset only. An example will be given in connection with the Hill estimator.

Recall that, by definition, (iv) is equivalent to

$$\frac{T(z_{\beta} + \lambda_{n}y_{n}) - T(z_{\beta})}{\lambda_{n}} \rightarrow T_{\beta}'(y)$$

for all $\lambda_{n} \rightarrow 0$ and all sequences $y_{n} \rightarrow y \in C_{\beta,h}$ where $T_{\beta}' : D_{\beta,h} \rightarrow \mathbb{R}$ has to be linear and continuous. By the Riesz representation theorem there exists a unique signed measure $v_{T,\beta}$ on $\mathcal{B}([0,1])$ such that

$$\int \frac{1}{h(t)\beta} |v_{T,\beta}|(dt) < \infty$$

(3.2)

and

$$T_{\beta}'(z) = \int z(t) \nu_{T,\beta}(dt)$$
for all $z \in C_{b,h}$. Let

$$
\sigma^2_{T,\beta} := \sigma^2_{b_{n},T,\beta} \quad \text{with} \quad \sigma^2_{b_{n},T,\beta} := e_\beta^2 \int (st)^{-(\beta+1)} \min(s, t) \, \nu^2(ds, dt)
$$

and

$$
\mu_{T,\beta} := e_\beta \int \bar{\Phi}(t) \, \nu_{T,\beta}(dt).
$$

Now we are ready for stating the main result of this section.

**Theorem 3.1.** Assume that $T$ satisfies Condition 3.

(i) Suppose that Condition 2(i) is satisfied and

Condition 1(i) or

Condition 1(ii) with $\beta = \delta$ or with $\beta < \delta$ and $d_2 \neq 0$ (cf. (3.1)).

Then

$$
\rho < \infty \quad \Rightarrow \quad \mathcal{L}(k_n^{1/2}(T(Q_n) - \beta)) \rightarrow \mathcal{N}(\rho \nu_{T,\beta}[0, 1], \sigma^2_{T,\beta}) \quad \text{weakly}
$$

and

$$
\rho = \infty \quad \Rightarrow \quad \mathcal{L}(k_n^{1/2}(T(Q_n) - \beta)) \rightarrow \nu_{T,\beta}[0, 1] \quad \text{in probability.}
$$

(ii) If Condition 1(ii) holds with $\beta > \delta$ or $\beta < \delta$ and $d_2 = 0$, then under Condition 2(ii)

$$
\lambda < \infty \quad \Rightarrow \quad \mathcal{L}(k_n^{1/2}(T(Q_n) - \beta)) \rightarrow \mathcal{N}(\lambda \mu_{T,\beta,\bar{\Phi}}, \sigma^2_{T,\beta}) \quad \text{weakly}
$$

and

$$
\lambda = \infty \quad \Rightarrow \quad \mathcal{L}(k_n^{1/2}(T(Q_n) - \beta)) \rightarrow \mu_{T,\beta,\bar{\Phi}} \quad \text{in probability.}
$$

Note that (3.2) guarantees that $\nu_{T,\beta}[0, 1]$, $\sigma^2_{T,\beta}$ and $\mu_{T,\beta,\bar{\Phi}}$ are well-defined real numbers.

**Proof.** If the second-order Condition 1(ii) holds with $\beta < \delta$, then by (3.1) one can choose $a(t) = d_1 \beta t^{-\beta}$, $c_\beta \Phi(t) = g(t) t^\beta / (d_1 \beta)$ and hence $\Xi(t) = d_2 / (d_1 t^\beta)$. In particular, $\Phi(t) = o(\Xi(t))$ as $t \downarrow 0$ if $d_2 \neq 0$ and $\Xi \equiv 0$ if $d_2 = 0$.

Condition 1(ii) with $\beta = \delta$ is equivalent to $g(t) := F^{-1}(1 - t) - dt^{-\beta} \in \Pi_0$ for some $d > 0$ with some auxiliary function $L$ (de Haan and Stadtmüller, 1996, Theorem 2). Since $L(t) = o(g(t))$ one obtains $c_\beta \Phi(t) = L(t) t^\beta / (d \beta)$ and $\Xi(t) \sim g(t) t^\beta / d$ and thus $\Phi(t) = o(\Xi(t))$.

Finally, if $\beta > \delta$, then for a suitable choice of the normalizing function $a$ one has $\Xi \equiv 0$.

To sum up, under the conditions of assertion (i) $\Xi$ dominates $\Phi$ and the converse holds true if the conditions of part (ii) are satisfied. Now the assertions can be proved in a similar way as Theorem 3.2 of Drees (1995b) by using the well-known $\delta$-method. $\square$
In general, the rate of convergence of statistical tail functionals $T$ with $v_{T,\beta}[0,1] \neq 0$ is slower than the rate for location invariant estimators if $\beta \leq \delta$. On the other hand, the rates are the same if $\beta > \delta$, but the variance $\sigma_{T,\beta}^2$ of the limit distribution (or its mean-squared error $\mu_{T,\beta,\psi}^2 + \sigma_{T,\beta}^2$) is smaller if a suitable functional $T$ is chosen that is not location invariant.

To see this, recall from Drees (1995b) that the regularity conditions on $T$ imply certain restrictions on the measure $v_{T,\beta}$. In the present situation, together with the location invariance, one of the conditions on $v_{T,\beta}$ is dropped, too.

**Lemma 3.1.** If $T$ satisfies Condition 3 for all $\beta$ of some open interval $I \subset (0, \infty)$, then

(i) $\int z_\beta \, dv_{T,\beta} = 0 \quad \forall \beta \in I$,

(ii) $\int z_\beta \cdot \log dv_{T,\beta} = -1 \quad \forall \beta \in I$.

Since the set of admissible measures is smaller if a third condition corresponding to the location invariance is added, one may expect that the minimal asymptotic variance is smaller in the present situation. In fact, whereas in the foregoing paper it was shown that the minimal asymptotic variance under location invariance equals $(\beta + 1)^2$, using the same methods one can easily see that

$$\inf_{v \in \mathcal{M}^*(\beta)} \sigma_{\beta,v}^2 = \beta^2,$$

where $\mathcal{M}^*(\beta)$ denotes the set of all signed measures satisfying (3.2) and the conditions given in Lemma 3.1(i) and (ii). Moreover, the infimum is attained at

$$v^*_{\beta} := z_{-\beta} \cdot \hat{\lambda}_{[0,1)} - \varepsilon_1,$$

where $\hat{\lambda}_{[0,1)}$ denotes the uniform distribution on $(0,1)$ and $\varepsilon_1$ the Dirac measure at 1.

Observe that $\beta^2$ is the asymptotic variance of the Hill estimator. Indeed in Reiss (1989), Section 9.4, it is proved that under more restrictive conditions the Hill estimator is asymptotically optimal if the bias is negligible. In the following example it is shown that the asymptotic normality of Hill’s estimator, which was already established in several papers before (see, e.g., Hall, 1982; Davis and Resnick, 1984; and Haeusler and Teugels, 1985), follows from Theorem 3.1.

**Example 3.1.** For any measurable function $z : [0, 1] \to \mathbb{R}$ let

$$T_H(z) := \int \log^+(z(t)/z(1)) \, dt$$

if the right-hand side is defined and finite and $T_H(z) = 0$ otherwise. Then $T_H(Q_n)$ is the Hill estimator.

Obviously, $T_H$ is scale invariant and $T_H(z_\beta) = \beta$. Now fix some $h \in \mathcal{H}$ that is increasing and regularly varying at 0 with exponent $\frac{1}{2}$ (e.g., $h(t) = (t/(1 - \log(t)))^{1/2}$) and define

$$\mathcal{D}_{\beta,h} := \{z \in D_{\beta,h} \mid z \text{ positive and nonincreasing}\}.$$
Since $P\{Q_n \in \tilde{D}_{\beta,h}\} = P\{X_n - k_n > 0\} \to 1$ it suffices to verify that $T_{H|\tilde{D}_{\beta,h}}$ is Hadamard differentiable at $z_\beta$ tangentially to $C_{\beta,h}$ (cf. the remark following Condition 3).

For this let $\lambda_n \to 0$ and $y_n \in D_{\beta,h}$ denote a sequence converging to some $y \in C_{\beta,h}$ such that $z_\beta + \lambda_n y_n \in \tilde{D}_{\beta,h}$ for all $n \in \mathbb{N}$. Then

$$T_H(z_\beta + \lambda_n y_n) - T_H(z_\beta) = \int_0^1 \log(1 + \lambda_n x_n(t)) \, dt,$$

where

$$x_n(t) := \frac{t^\beta y_n(t) - y_n(1)}{1 + \lambda_n y_n(1)}.$$

The monotonicity of $z_\beta + \lambda_n y_n$ implies $\log(1 + \lambda_n x_n(t)) \geq -\beta \log(s/t) + \log(1 + \lambda_n x_n(s))$ for all $t \leq s$ and hence,

$$\int_0^s \log(1 + \lambda_n x_n(t)) \, dt \geq s(-\beta + \log(1 + \lambda_n x_n(s))). \quad (3.3)$$

Moreover, for $s_n = 7/6$ one has

$$\sup_{t \in [s_n, 1]} \log(1 + \lambda_n x_n(t)) \leq \lambda_n \sup_{t \in [s_n, 1]} |x_n(t)| = O(\lambda_n/h(s_n)) = o(1). \quad (3.4)$$

Using (3.3), (3.4) and $\log(1 + x) \leq x$ one obtains

$$\left| \int_{s_n}^{s_n} \log(1 + \lambda_n x_n(t)) - \lambda_n x_n(t) \, dt \right| = O(s_n) = o(\lambda_n).$$

Furthermore, a combination of (3.4) and $\log(1 + x) = x + O(x^2)$ as $x \to 0$ shows that

$$\left| \int_{s_n}^{1} \log(1 + \lambda_n x_n(t)) - \lambda_n x_n(t) \, dt \right| = O\left( \int_{s_n}^{1} (\lambda_n x_n(t))^2 \, dt \right)$$

$$= O\left( \lambda_n^2 \sup_{t \in [s_n, 1]} |x_n(t)| \int_0^1 1/h(t) \, dt \right)$$

$$= o(\lambda_n).$$

To sum up, we have proved that

$$\left| \int_0^1 \log(1 + \lambda_n x_n(t)) - \lambda_n x_n(t) \, dt \right| = o(\lambda_n),$$

which in turn is equivalent to the Hadamard differentiability of $T_{H|\tilde{D}_{\beta,h}}$ at $z_\beta$ with $T_{H,\beta}'(y) = \int_0^1 t^\beta y(t) - y(1) \, dt$, because obviously $|\int_0^1 x_n(t) \, dt - T_{H,\beta}'(y)| = o(1)$. Notice that the signed measure pertaining to $T_{H,\beta}'$ is the one that minimizes the asymptotic variance.

In a similar way, one may check that the moment estimator proposed by Dekkers et al. (1989) and kernel-type estimators (Csörgő et al., 1985) satisfy Condition 3 and
are therefore asymptotically normal. This was already proved in the above-mentioned papers under comparable second-order conditions.

Furthermore, Theorem 3.1 allows the construction of statistical tail functionals with prescribed asymptotic behavior.

**Example 3.2.** In this example we want to design a functional $T$ satisfying Condition 3 for all $\beta > 0$ such that the signed measures corresponding to the Hadamard derivatives equal a given family of measures $v_\beta$. Of course, it has to be assumed that $v_\beta$ satisfies (3.2) and the conditions given in Lemma 3.1. Moreover, it is plausible that the family $(v_\beta)_{\beta > 0}$ should be smooth in some sense. More precisely, we assume that for all $\beta_n \to \beta$

$$\int \frac{1}{h(t)^{\beta_n}} |v_{\beta_n} - v_{\beta}|(dt) \to 0. \quad (3.5)$$

In particular, for some $\varepsilon > 0$

$$\int t^{-(\beta+1/2+\varepsilon)} |v_{\beta}|(dt) < \infty. \quad (3.6)$$

For instance, the continuity Condition (3.5) is fulfilled by the family of measures $v_\beta$ that leads to a minimal asymptotic variance.

We proceed as follows: first we construct a functional with the given Hadamard derivative for a fixed $\beta > 0$ and then an adaptive version of this functional is introduced. For simplicity, it is assumed that $\lim_{t \to 0} h(t) t^{-(1/2+\varepsilon)} = \infty$ for all $\varepsilon > 0$.

Let

$$H_\beta(t) := t^\beta \int_{[0,t]} s^{-\beta} v_\beta(ds)$$

and define a functional

$$T_\beta(z) := \frac{\int z(t) dH_\beta(t)}{\int z(t) H_\beta(t)/t \, dt}$$

(with the convention $x/0 = 0$ for all $x \in \mathbb{R}$). Note that $T_\beta(Q_n)$ is a generalized probability weighted moment estimator (cf. Drees, 1995b, Example 1.1).

Integration by parts yields

$$dH_\beta(t) = v_\beta(dt) + (\beta H_\beta(t)/t) \, dt \quad (3.7)$$

and a combination with (3.6) shows that $T_\beta$ is well defined. Obviously, it is scale invariant and because of Lemma 3.1(i) Condition 3(iii) is also satisfied. Finally,

$$\int z_\beta(t) H_\beta(t)/t \, dt = - \int \log(t) z_\beta(t) v_\beta(dt) = 1 \quad (3.8)$$

and hence it is readily verified that for $\lambda_n \to 0$ and $y_n \to y \in C_{\beta,h}$

$$T_\beta(z_\beta + \lambda_n y_n) - \beta = \frac{\lambda_n \int y_n(t) v_\beta(dt)}{1 + \lambda_n \int y_n(t) H_\beta(t)/t \, dt} = \hat{\lambda}_n \int y(t) v_\beta(dt) + o(\lambda_n),$$
that is, $T_\beta$ is Hadamard differentiable at $z_\beta$ with measure $v_\beta$ pertaining to its derivative.

Now define an "adaptive functional" by

$$T(z) := T_\tilde{\beta}(z)$$

if the right-hand side is well defined and $T(z) := 0$ else. Here $\tilde{\beta}$ is some functional that is continuous at $z_\beta$ and satisfies the Conditions 3(i)–(iii) for all $\beta > 0$ (e.g., the Hill functional $T_H$) so that $\tilde{\beta}(Q_n)$ is a consistent initial estimate of $\beta$.

Obviously, $T$ fulfills Condition 3(ii) and (iii) and it remains to prove (iv). To this end, consider sequences $\lambda_n \to 0$ and $y_n \to y \in C_{\beta,h}$ and let $\beta(n) := \tilde{\beta}(z_\beta + \lambda_n y_n)$. Since $\beta(n) \to \beta$, one has $\lim_{t \to 0} t^{-\beta} H_{\beta(n)}(t) = 0$ for sufficiently large $n$ and thus using integration by parts one obtains

$$T_\beta(n)(z_\beta) = \beta.$$

Next, observe that the continuity property (3.5) is passed on from $(v_\beta)_{\beta > 0}$ to $(dH_\beta)_{\beta > 0}$. Consequently,

$$T(z_\beta + \lambda_n y_n) - \beta = \frac{\int z_\beta(t) + \lambda_n y_n(t) dH_{\beta(n)}(t)}{\int (z_\beta(t) + \lambda_n y_n(t))H_{\beta(n)}(t)/t \, dt} - \frac{\int z_\beta(t) dH_{\beta(n)}(t)}{\int z_\beta(t)H_{\beta(n)}(t)/t \, dt} + o(\lambda_n)$$

$$= \lambda_n \int y(t) v_\beta(dt) + o(\lambda_n),$$

where for the last equation (3.7), (3.8) and 3.1(i) have been utilized. Hence, $T$ is Hadamard differentiable at $z_\beta$ with the preassigned derivatives for all $\beta > 0$.

It should be mentioned that, in a similar way, one can construct statistical tail functionals with given Hadamard derivatives that are also location invariant.

Example 3.2 describes a method for designing statistical tail functionals with made-to-order asymptotic behavior. If one only considers the asymptotic variance, then such a construction is superfluous since there is a simple estimator, namely, the Hill estimator, with minimal asymptotic variance. Yet if the bias is taken into account, i.e., if one considers the asymptotic mean-squared error (MSE), then Hill’s estimator is not optimal (cf. Smith, 1987, Section 4). In fact, there does not exist a statistical tail functional that has a minimal asymptotic MSE simultaneously for all underlying d.f.’s $F$ satisfying Condition 1(ii) (cf. Drees, 1995b, Section 4). The best one can do is to use a functional such that the leading term of the bias vanishes if the underlying d.f. $F$ satisfies Condition 1(ii) with $\delta > 0$ belonging to a given finite set. For this again one has to distinguish the two cases of Theorem 3.1.

In the situation of Theorem 3.1(i) $T_\beta(E(k_n/n))$ is the dominating part of the bias of the estimator. So if one chooses $T$ in such a way that this term vanishes asymptotically,
then $T$ has to be ‘almost location invariant’ locally at $z_{\beta}$. Essentially, in this case we are in the situation of Drees (1997) since $T'_\beta(\Xi(k_n/n)) = 0$ implies $\nu_{T,\beta}[0,1] = 0$, which is exactly the additional condition that follows from location invariance.

Therefore, in the following, we assume that the conditions of Theorem 3.1(ii) are fulfilled. Using Example 3.2, for each positive $\delta_0$ one may construct functionals $T$ such that the bias term $T'_\beta(\Psi)$ vanishes asymptotically if the second-order Condition 1(ii) is satisfied for $\delta = \delta_0$. Within this class of functionals the minimal asymptotic variance $\sigma_{\beta,v}^2$ equals $((\delta_0 + 1)/\delta_0)^2 \beta^2$ and it is attained for

$$\nu_{\beta,\delta_0}(dt) = \frac{\delta_0 + 1}{\delta_0} \beta \left( \frac{\delta_0 + 1}{\delta_0} (\tau^{\beta} - (2\delta_0 + 1)R^{\beta}) dt - \varepsilon_1(dt) \right).$$

The situation is quite different if Condition 1(ii) holds with $\delta = 0$, because then the asymptotic MSE of all statistical tail functionals is the same if the number of upper order statistics used for estimation is chosen optimally. So in this case all statistical tail functionals have the same asymptotic efficiency. For a proof of these results and an extensive discussion of the choice of $T$ if the bias is taken into account we refer to Drees (1995b), Section 4.

4. Asymptotics for $\beta \leq 0$

If $\beta < 0$ and the right endpoint $\omega$ of the underlying d.f. equals 0, then $\Xi = 0$, and consequently, one can establish asymptotic normality by following the lines of Section 3. In contrast to that, if $\beta = 0$ or $\beta < 0$ and $\omega \neq 0$, then $\lim_{t \to 0} \Xi(t) = \infty$ so that no longer $z_{\beta}$ but $\Xi(k_n/n)$ is the dominating term in the approximation of $Q_n$. Therefore, the smoothness condition on the functional $T$ must determine its local behavior in a neighborhood of a constant function instead of $z_{\beta}$.

Since for $\beta < -\frac{1}{2}$, in general, $D_{\beta,h}$ does not include the constant functions, $T$ has to be defined on $\text{span}(D_{\beta,h},1)$. This function space will be equipped with the $\sigma$-field $\mathcal{A}_{\beta,h}$ that is generated by all sets of the form $\{z + c \mid z \in B, c \in [c_1,c_2]\}$ with $B \in \mathcal{B}(D_{\beta,h})$ and $-\infty < c_1 < c_2 < \infty$. For the sake of definiteness, in the sequel we will assume that $\omega > 0$, but the case $\omega < 0$ can be treated in the same way.

**Condition 4.** Assume that $T : \text{span}(D_{\beta,h},1) \to \mathbb{R}$ satisfies for $\beta \leq 0$ and some $h \in \mathcal{H}$

(i) $T$ is $\mathcal{A}_{\beta,h},\mathcal{B}(\mathbb{R})$-measurable,

(ii) $T(az) = T(z) \quad \forall z \in \text{span}(D_{\beta,h},1), \quad a > 0,$

(iii) $T(1 + \rho_n(z_{\beta} + \lambda_n y_n)) = \beta + \rho_n c_{T,\beta} + \lambda_n T'_{\beta}(y) + o(\rho_n + \lambda_n)$ for all $\rho_n, \lambda_n \downarrow 0$ and $y_n \to y \in C_{\beta,h}$ with $T'_{\beta} : D_{\beta,h} \to \mathbb{R}$ denoting some continuous, linear functional.

Note that despite the notation $T'_{\beta}$ is not a proper derivative. At first glance, Condition 4(iii) seems rather peculiar but Theorem 2.1 suggests that, nevertheless, it is appropriate in the present situation. Below we will give two examples of such functionals including the moment estimator due to Dekkers et al. (1989). It is worth mentioning that there are straightforward generalizations of Condition 4(iii) that lead to similar results on the
asymptotic behavior of $T(Q_n)$. For example, on the right-hand side of the expansion, one may replace $\rho_n$ with $g(\rho_n)$ where $g(\rho)$ is some function converging to 0 as $t$ tends to 0.

In view of Condition 4(iii) for the convergence rate of $T(Q_n)$ the rates of $\varepsilon(k_n/n)^{-1}$ and $\Phi(k_n/n)$ are decisive. In analogy to the case $\beta>0$, $\varepsilon(k_n/n)^{-1}$ is dominating if $|\beta|<\delta$ and $\varepsilon(k_n/n)^{-1}=o(\Phi(k_n/n))$ if $|\beta|>\delta$ but there is no plain answer in the case $|\beta|=\delta$. So to keep the conditions on $k_\alpha$ as simple as possible we restrict ourselves to the case of a nondegenerate limiting distribution of the standardized estimator and leave the other cases to the reader, though this excludes the optimal choice of $k_\alpha$ if $\beta=0$ or $\delta=0$.

**Condition 5.** Let $(k_n)_{n \in \mathbb{N}}$ be some intermediate sequence satisfying

(i) $k_n^{1/2} \sup_{x \in (0,1+\varepsilon)} x^{\beta+1/2} |R(k_n/n,x)| \to 0$,

(ii) $k_n^{1/2}/\varepsilon(k_n/n) \to \rho \in [0, \infty)$,

(iii) $k_n^{1/2} \Phi(k_n/n) \to \lambda \in [0, \infty)$.

Using the notation of Section 3 we have the following limit theorem.

**Theorem 4.1.** Suppose $T$ satisfies Condition 4.

(i) If Condition 1(i) holds or (ii) with $|\beta|<\delta$, then under Conditions 5(i) and (ii)

$$\mathcal{L}(k_n^{1/2}(T(Q_n)-\beta)) \to \mathcal{N}(\rho c_{T,\beta}, \sigma_{T,\beta}^2) \text{ weakly},$$

(ii) Under Condition 1(ii) with $|\beta|>\delta$ and 5(iii) one has

$$\mathcal{L}(k_n^{1/2}(T(Q_n)-\beta)) \to \mathcal{N}(\lambda \mu_{T,\beta}, \psi, \sigma_{T,\beta}^2) \text{ weakly},$$

(iii) Condition 1(ii) with $|\beta|=\delta$ in combination with Conditions 5(ii) and (iii) implies

$$\mathcal{L}(k_n^{1/2}(T(Q_n)-\beta)) \to \mathcal{N}(\rho c_{T,\beta} + \lambda \mu_{T,\beta}, \psi, \sigma_{T,\beta}^2) \text{ weakly}.$$

**Proof.** One has to examine the cases considered in Theorem 2 of de Haan and Stadtmüller (1996) separately. If $\beta=0$ and $\delta>0$, then $\varepsilon(t) \sim -\log(t)$ and hence $\Phi(t)=o(\varepsilon(t)^{-1})$. If $\beta<0$, then $\varepsilon(t)^{-1}$ is $|\beta|$-varying at 0 so that $\Phi(t)=o(\varepsilon(t)^{-1})$ if $|\beta|<\delta$ and $\varepsilon(t)^{-1}=o(\Phi(t))$ if $|\beta|>\delta$. Moreover, $\Delta_n=O_p(a(1/n)/a(k_n/n))=o_p(k_n^{1/2})$ for $\beta<\frac{1}{2}$ because $a$ is $|\beta|$-varying at 0.

Thus, according to the Skorohod representation theorem under the conditions of the theorem there are versions of $Q_\alpha$ and $V_\beta$ such that almost surely

$$Q_n/c_n \varepsilon(k_n/n) = 1 + \varepsilon(k_n/n)^{-1}(z_\beta + k_n^{-1/2}(\lambda \bar{\psi}_\beta + V_\beta + o(1))),$$

with $\lambda=0$ if $|\beta|<\delta$ or merely Condition 1(i) is assumed. Consequently, by Condition 4

$$T(Q_n) = \beta + \varepsilon(k_n/n)^{-1} c_{T,\beta} + k_n^{-1/2} T_\beta(\lambda \bar{\psi}_\beta + V_\beta) + o(\varepsilon(k_n/n)^{-1} + k_n^{-1/2}),$$

which in turn implies the assertion.
Example 4.1. For a measurable function $z \in [0, 1] \rightarrow \mathbb{R}$ let

$$T_M(z) := T_1(z) + 1 - \left( 2 \left( 1 - \frac{T_1(z)^2}{T_2(z)} \right) \right)^{-1}$$

with

$$T_i(z) := \int_0^1 (\log^+ (z(t)/z(1)))^i \, dt$$

if the right-hand side is well defined and finite and $T_M(z) := 0$ else. Then $T_M(Q_n)$ is the moment estimator introduced by Dekkers et al. (1989).

Obviously $T_M$ is scale invariant. Next we give a sketch of a proof of Condition 4(iii) under the additional assumption that the supremum of $y_n$ is bounded, which holds always true if $\beta < -\frac{1}{2}$ and $h \in \mathcal{H}$ is chosen appropriately. The general case can be treated in a similar way as Example 3.1.

A Taylor expansion of the logarithm yields

$$T_1(1 + \rho_n(z_\beta + \lambda_n y_n))$$

$$= \rho_n \left( \int_0^1 z_\beta(t) - z_\beta(1) \, dt + \lambda_n \int_0^1 y(t) - y(1) \, dt ight.$$

$$- \frac{\rho_n}{2} \int_0^1 z_\beta^2(t) - z_\beta^2(1) \, dt + o(\rho_n + \lambda_n) \right)$$

and

$$T_2(1 + \rho_n(z_\beta + \lambda_n y_n))$$

$$= \rho_n^2 \left( \int_0^1 (z_\beta(t) - z_\beta(1))^2 + 2\lambda_n \int_0^1 (z_\beta(t) - z_\beta(1)) \right.$$

$$\times (y(t) - y(1)) \, dt - \rho_n \int_0^1 (z_\beta(t) - z_\beta(1))(z_\beta^2(t) - z_\beta^2(1)) \, dt + o(\rho_n + \lambda_n) \right).$$

Straightforward but lengthy computations show that, for $\beta < 0$,

$$T_M(1 + \rho_n(z_\beta + \lambda_n y_n)) = \beta + \rho_n \frac{\beta(2 - 5\beta + \beta^2)}{(1 - \beta)(1 - 3\beta)} + \frac{\lambda_n (1 - \beta)(1 - 2\beta)}{\beta}$$

$$\times \left( \frac{1 - \beta}{\beta} \int_0^1 (1 - (1 - 2\beta)t^{-\beta})z(t) \, dt - z(1) \right)$$

$$+ o(\rho_n + \lambda_n)$$

and

$$T_M(1 + \rho_n(z_0 + \lambda_n y_n)) = 2\rho_n + \lambda_n \left( \int_0^1 (2 + \log(t))z(t) \, dt - z(1) \right) + o(\rho_n + \lambda_n).$$
Hence, Condition 4(iii) is satisfied and
\[ v_{\beta}(dt) = \begin{cases} \frac{(1-\beta)(1-2\beta)}{\beta} \left( \frac{1-\beta}{\beta} (1 - (1 - 2\beta)t^{-\beta}) dt - \varepsilon_1(dt) \right) & \beta < 0, \\ (2 + \log(t)) dt - \varepsilon_1(dt) & \beta = 0. \end{cases} \]

By Theorem 4.1 the asymptotic normality of the moment estimator follows under slightly more general assumptions than Dekkers et al. (1989) used in their paper.

Note that \( v_{\beta,0} \) is the signed measure that minimizes the asymptotic variance if \( \beta = 0 \) and location invariance is assumed (Drees, 1995b, Theorem 4.1). Generally, \( v_{\beta,0} \) satisfies the conditions given in Lemma 4.1 of Drees (1995b) for location invariant statistical tail functionals, i.e., the conditions of Lemma 3.1 and, in addition, \( v_{\beta,0}[0,1] = 0 \) if \( \beta \geq -\frac{1}{2} \). Later on we will see that under a very weak extra condition this always holds true for functionals satisfying Condition 4.

The following example indicates a general method to construct functionals satisfying Condition 4.

Example 4.2. Let \( T_i, i = 1,2, \) be scale invariant measurable functionals satisfying the expansion
\[ T_i(1 + \lambda z) = \lambda_n T'_i(z) + \lambda_n^2 \tilde{T}_i(z) + o(\lambda_n^2) \]
uniformly for \( \|z\|_{\beta,h} \leq 1 \) where \( T'_i \) is a continuous and linear and \( \tilde{T}_i \) some continuous functional on \( D_{\beta,h} \). (A typical example of such a functional is \( T(z) = \int f(z(t)/z(1)) \mu(dt) \) where \( f \) is some sufficiently smooth function with \( f(1) = 0 \) and \( \mu \) is some signed measure on \( \mathcal{B}([0,1]) \) that does not concentrate too much mass in a neighborhood of 0.)

If one has \( T'_i(z_{\beta})/T'_2(z_{\beta}) = \beta \), then it is easily seen that
\[ T(z) := T_1(z)/T_2(z) \]
satisfies Condition 4 with
\[ c_{T,\beta} = \frac{\tilde{T}_1(z_{\beta})T'_2(z_{\beta}) - \tilde{T}_2(z_{\beta})T'_1(z_{\beta})}{T'_2(z_{\beta})^2} \]
and
\[ T'_\beta(y) = \frac{T'_1(y)T'_2(z_{\beta}) - T'_2(y)T'_1(z_{\beta})}{T'_2(z_{\beta})^2}. \]

According to Theorem 4.1 the rate of convergence of estimators \( T(Q_n) \) with \( T \) satisfying Condition 4 and \( c_{T,\beta} \neq 0 \) is slower than the rate for location invariant estimation functionals if Condition 1(ii) holds with \( |\beta| < \delta \). This disadvantage is most clearly demonstrated by an exponential sample where any rate of convergence slower than \( n^{-1/2} \) is achieved by a smooth location invariant statistical functional based on a suitably chosen number of order statistics whereas under Condition 4 with \( c_{T,0} \neq 0 \) the best
attainable rate is $1/\log(n)$. Yet, in contrast to the situation examined in Section 3, the following lemma shows that asymptotically there is no advantage in using functionals that are not location invariant in the case $|\beta| > \delta$ either.

**Lemma 4.1.** If $T$ satisfies Condition 4, then

(i) $\int z_\beta \, dv_{T,\beta} = 0$.

(ii) $v_{T,\beta}[0,1] = 0$.

(iii) If, in addition,

$$T(1 + \rho_n z_{\beta_n}) = \beta_n + \rho_n c_{T,\beta_n} + o(\rho_n)$$

for all $\rho_n \downarrow 0$ and $\beta_n - \beta = O(\rho_n)$, then one has

$$\int z_\beta \log dv_{T,\beta} = \begin{cases} -1 & \beta \neq 0, \\ -2 & \beta = 0. \end{cases}$$

**Proof.** By Conditions 4(ii) and (iii) one has

$$\beta + \rho_n c_{T,\beta} + o(\rho_n) = T(1 + \rho_n z_{\beta}) = T(1 + \rho_n(1 + \rho_n)(z_\beta \rho_n - (1 + \rho_n))) = \beta + \rho_n c_{T,\beta} + \rho_n T'_n(1) + o(\rho_n)$$

and hence $T'_n(1) = 0$, i.e., assertion (ii). In a similar way, Condition 4(iii) applied to $\lambda_n y_n = \rho_n z_{\beta}$ yields (i).

Now assume that (4.1) holds and $\beta < 0$. Then $(z_{\beta_n} - z_\beta)/(\beta_n - \beta) \rightarrow -z_\beta \log$ in $D_{\beta,\delta}$ and hence an application of 4(iii) to $\lambda_n y_n = z_{\beta_n} - z_\beta$ gives

$$(\beta_n - \beta)(1 + T'_n(z_\beta \log)) + \rho_n (c_{T,\beta_n} - c_{T,\beta}) = o(\rho_n)$$

for all $\beta_n - \beta = O(\rho_n)$. So the choice $\beta_n - \beta = o(\rho_n)$ shows that the map $\beta \mapsto c_{T,\beta}$ has to be continuous and thus

$$(\beta_n - \beta)(1 + T'_n(z_\beta \log)) = o(\rho_n)$$

for all $\beta_n - \beta = O(\rho_n)$, which in turn implies (iii).

Likewise, assertion (iii) can be proved in the case $\beta = 0$ by utilizing $2\beta_n^{-2}(z_{\beta_n} - 1 + \beta_n z_0) \rightarrow -z_0 \log$. $\Box$

The additional condition (4.1) is very mild since, roughly speaking, it means that in case $y_n \equiv 0$ the Condition 4 holds locally uniformly for $\beta \leq 0$. Note that for $\beta \geq -\frac{1}{2}$ the conditions given in Lemma 4.1 (i)–(iii) are equivalent to the conditions established in Drees (1995b), Lemma 4.1, for location invariant functionals, while in the case $\beta < -\frac{1}{2}$ they are even stronger. Hence, for each functional satisfying Condition 4 there exists a location invariant statistical tail functional with the same limit distribution if $|\beta| > \delta$ and a faster rate of convergence towards the true parameter if $|\beta| < \delta$. In view
of this result we strongly recommend to use statistical tail functionals that are location invariant if one expects $\beta$ to be nonpositive.

Acknowledgements

Part of the work was completed while the author was visiting the Erasmus University Rotterdam, supported by a DFG research grant. The author would like to thank L. de Haan for his hospitality.

References